数理解析研究所講究録
1114 巻 1999 年 22-32

A fuzzy treatment of uncertain Markov decision processes

千葉大学教育学部 蔵野正美 (Masami KURANO)
千葉大学理学部 安田正實 (Masami YASUDA)
千葉大学理学部 中神潤一 (Jun-ichi NAKAGAMI)
北九州大学経済学部 吉田祐治 (Yuji YOSHIDA)

Abstract

In this paper, we shall develop a fuzzy treatment for uncertain Markov decision processes which allow for a fluctuating transition matrix at each step in time. The decision model with uncertain transition matrices is described by the use of fuzzy sets in which we find a Pareto optimal policy maximizing the infinite horizon fuzzy expected discounted reward over all stationary policies under some partial order. The Pareto optimal policies are characterized by maximal solutions of an optimal equation including efficient set-functions. As a numerical example, the machine maintenance problem is considered.

Keywords: Fuzzy analysis, Markov decision process, Pareto optimal, optimality equation.
AMS 1991 subject classification. Primary: 90c40; Secondary: 90c39.

1. Introduction

In a real application of Markov decision processes (MDP, in short, cf.[1, 5, 7, 13, 16]), we often encounter the case where the required data is not known precisely and perfectly. In fact, in many instances, the required data in MDPs must be estimated through the measurement of various phenomena, so that it naturally includes imprecision or ambiguity of the observing system. Also, it requires to be more "robust" in the sense that it is reasonably efficient in approximations.

In order to deal with these uncertain data and flexible requirements, Kruce et al.[8] have used a fuzzy set representation for homogeneous Markov chains with uncertain transition matrices, in which ergodic theorems are obtained in fuzzy environment.

In this paper, we shall develop a fuzzy treatment for uncertain MDPs which allow for fluctuating transition matrices at each step in time. The MDPs with uncertain transition matrices are described by the use of fuzzy sets, in which we find a Pareto optimal policy maximizing the infinite horizon fuzzy expected discounted reward(FEDR) over all stationary policies under some partial order relation.

Associated with each stationary policy is a corresponding contractive operator on fuzzy numbers, whose fixed point represents the infinite horizon FEDR. Moreover, the Pareto optimal policies are characterized by maximal solutions of an optimal equation including efficient set-functions. As a numerical example, the machine maintenance problem is considered.

Recently, applying Hartfiel's[3, 4] interval method for Markov chains, Kurano et al.[10] have introduced a decision model, called a controlled Markov set-chain, which is robust for rough approximation of transition matrices in MDPs.

Our fuzzy decision model examined in this paper includes a controlled Markov set-chain as a special case. So, the results obtained here can be thought of as a fuzzy extension of those in [10]. For the optimization of fuzzy dynamic system, refer to [9, 18]. The non-discounted reward problem for a controlled Markov set-chain was developed in [6, 11].

This paper is organized as follows: In Section 2, we shall give some notation on fuzzy sets and interval arithmetics and obtain the preliminary lemmas. In Section 3, we describe a nonhomogeneous MDPs by the use of fuzzy sets and specify the optimization problem. In Section 4, the infinite horizon FEDR from a stationary policy is given as a fixed point of a corresponding operator, which is used to obtain the optimality equation and characterize a Pareto optimal policy in Section 5.

2. Notation and preliminary lemmas

Let $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$ be set of real numbers, real n-dimensional column vectors and real $n \times n$ matrices, respectively. Also denote by $\mathbb{R}_+$, $\mathbb{R}^n_+$ and $\mathbb{R}^{n\times n}_+$, the subsets of entrywise non-negative elements in $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$, respectively. We provide $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$ with the componentwise relation $\le$ and $<$. For any set X, we will denote a fuzzy set $\widetilde{a}$ on X by its membership function $\widetilde{a} : X \to [0,1]$. Denote by $\mathcal{F}(X)$ the set of all fuzzy sets on X. For the theory of fuzzy sets, refer to Zadeh[19] and Novák[15]. The α-cut ($\alpha \in [0,1]$) of the fuzzy set $\widetilde{a} \in \mathcal{F}(X)$ is defined as

$$\widetilde{a}_\alpha := \{x \in X \mid \widetilde{a}(x) \ge \alpha\} \ (\alpha > 0) \quad \text{and} \quad \widetilde{a}_0 := \mathrm{cl}\{x \in X \mid \widetilde{a}(x) > 0\},$$

where cl denote the closure of the set. For any interval Y in $\mathbb{R}$, $\widetilde{a} \in \mathcal{F}(Y)$ is called a fuzzy number on Y if $\widetilde{a}$ has the following properties (i) – (iv):

(i) $\widetilde{a}$ is normal, i.e., there exists an $x_0 \in Y$ with $\widetilde{a}(x_0) = 1$;

(ii) $\widetilde{a}$ is convex, i.e., $\widetilde{a}(\alpha x + (1-\alpha)y) \ge \widetilde{a}(x) \wedge \widetilde{a}(y)$ for all $x, y \in Y$ and $\alpha \in [0,1]$, where $a \wedge b = \min\{a, b\}$;

(iii) $\widetilde{a}$ is upper semi-continuous;

(iv) $\widetilde{a}_0$ is a compact subset of Y.

Denote by $\mathcal{F}_c(Y)$ the set of all fuzzy numbers on Y. Let $\mathcal{C}(Y)$ be the set of all closed and bounded intervals in Y. We note that $\widetilde{a} \in \mathcal{F}_c(Y)$ means $\widetilde{a}_\alpha \in \mathcal{C}(Y)$ for all $\alpha \in [0,1]$. Let $\mathcal{F}_c(Y)^n$ be the set of all n-dimensional column vectors whose elements are in $\mathcal{F}_c(Y)$, i.e.,

$$\mathcal{F}_c(Y)^n := \{\widetilde{u} = (\widetilde{u}_1, \widetilde{u}_2, \dots, \widetilde{u}_n)' \mid \widetilde{u}_i \in \mathcal{F}_c(Y) \ (1 \le i \le n)\},$$

where d' denotes the transpose of a vector d.

Let $S := \{1, 2, \dots, n\}$ and $\mathcal{P}(S)$ the set of all probability distributions on S, that is,

$$\mathcal{P}(S) := \{p = (p_1, p_2, \dots, p_n) \mid p_j \ge 0 \ (1 \le j \le n), \sum_{j=1}^{n} p_j = 1\}.$$

From any $\widetilde{p} = (\widetilde{p}_1, \widetilde{p}_2, \dots, \widetilde{p}_n)' \in \mathcal{F}_c([0,1])^n$, we will construct the fuzzy set $[\widetilde{p}] = [\widetilde{p}_1, \widetilde{p}_2, \dots, \widetilde{p}_n]$ on $\mathcal{P}(S)$ by the following:

$$[\widetilde{p}](p) = \min_{1\le j\le n} \{\widetilde{p}_j(p_j)\} \quad \text{for any} \quad p = (p_1, p_2, \dots, p_n) \in \mathcal{P}(S). \tag{2.1}$$

The above definition will be extended to the case of stochastic matrices. Let $\mathcal{P}(S/S)$ be the set of all stochastic matrices on S, that is,

$$\mathcal{P}(S/S) := \{Q = (q_{ij}) \mid q_{ij} \ge 0, \sum_{j=1}^{n} q_{ij} = 1 \ (1 \le i \le n)\}.$$

For any $\widetilde{q}_i = (\widetilde{q}_{i1}, \widetilde{q}_{i2}, \ldots, \widetilde{q}_{in}) \in \mathcal{F}_c([0,1])^n$ $(1 \le i \le n)$, we define the fuzzy set $\widetilde{Q} = [\widetilde{q}_1, \widetilde{q}_2, \ldots, \widetilde{q}_n]'$ on $\mathcal{P}(S/S)$ as follows:

$$\widetilde{Q}(Q) := \min_{1\le i\le n} \{[\widetilde{q}_i](q_i)\}, \tag{2.2}$$

where $Q = (q_1, q_2, \ldots, q_n)' \in \mathcal{P}(S/S)$, $q_i = (q_{i1}, q_{i2}, \ldots, q_{in}) \in \mathcal{P}(S)$ and $[\widetilde{q}_i]$ is the fuzzy set on $\mathcal{P}(S)$ defined by (2.1).

In order to describe the structural properties on the fuzzy sets defined in (2.1) and (2.2), we need the concept of intervals of matrices. For the detail, refer to [4, 10, 14]. For any nonnegative vector $\underline{q} = (\underline{q}_1, \underline{q}_2, \ldots, \underline{q}_n)$ and $\overline{q} = (\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_n) \in \mathbb{R}^n_+$ with $\underline{q} \le \overline{q}$, we define the interval $\langle \underline{q}, \overline{q}\rangle \subset \mathcal{P}(S)$ by

$$\langle \underline{q}, \overline{q}\rangle := \{p = (p_1, p_2, \ldots, p_n) \in \mathcal{P}(S) \mid \underline{q} \le p \le \overline{q}\}. \tag{2.3}$$

Similarly, for $\underline{Q} = (\underline{q}_{ij}), \overline{Q} = (\overline{q}_{ij}) \in \mathbb{R}^{n\times n}_+$ with $\underline{Q} \le \overline{Q}$,

$$\langle \underline{Q}, \overline{Q}\rangle := \{Q \in \mathcal{P}(S/S) \mid \underline{Q} \le Q \le \overline{Q}\}. \tag{2.4}$$

Lemma 2.1 ([4]). *For any $\underline{Q}, \overline{Q} \in \mathbb{R}^{n\times n}_+$ with $\underline{Q} \le \overline{Q}$ and $\langle \underline{Q}, \overline{Q}\rangle \ne \emptyset$, $\langle \underline{Q}, \overline{Q}\rangle$ is a polyhedral convex set in the vector space $\mathbb{R}^{n\times n}$.*

For any $\widetilde{a} \in \mathcal{F}_c([0,1])$, noting $\widetilde{a}_\alpha \in \mathcal{C}([0,1])$ $(0 \le \alpha \le 1)$, it will be denoted by $\widetilde{a}_\alpha = [\min \widetilde{a}_\alpha, \max \widetilde{a}_\alpha]$. The structural property of the fuzzy sets defined in (2.1) and (2.2) is given, whose proof is done by using Lemma 2.1.

Lemma 2.2. *For any $\widetilde{q}_i \in \mathcal{F}_c([0,1])^n$ $(1 \le i \le n)$, let $\widetilde{Q} = [\widetilde{q}_1, \widetilde{q}_2, \ldots, \widetilde{q}_n]'$ be a fuzzy set on $\mathcal{P}(S/S)$ defined by (1.2). Then, the α-cut of $\widetilde{Q}$ $(0 \le \alpha \le 1)$ is a polyhedral convex subset of $\mathcal{P}(S/S)$ and given by*

$$\widetilde{Q}_\alpha = \langle \underline{Q}_\alpha, \overline{Q}_\alpha\rangle, \quad \textit{where } \underline{Q}_\alpha = \big(\min(\widetilde{q}_{ij})_\alpha\big) \textit{ and } \overline{Q}_\alpha = \big(\max(\widetilde{q}_{ij})_\alpha\big). \tag{2.5}$$

Proof. Since $\widetilde{q}_{ij} \in \mathcal{F}_c([0,1])$, the α-cut $(\widetilde{q}_{ij})_\alpha$ belongs to $\mathcal{C}([0,1])$. By (2.1) and (2.2), we observe that

$$\widetilde{Q}_\alpha = \{Q = (q_{ij}) \in \mathcal{P}(S/S) \mid q_{ij} \in (\widetilde{q}_{ij})_\alpha \ (1 \le i, j \le n)\},$$

which implies that (2.5) holds. Thus, by Lemma 2.1, $\widetilde{Q}_\alpha$ has the required property. □

If $\boldsymbol{u} = ([a_1, b_1], [a_2, b_2], \ldots, [a_n, b_n])' \in \mathcal{C}(\mathbb{R}_+)^n$, $\boldsymbol{u}$ will be denoted by $\boldsymbol{u} = [a, b]$, where $a = (a_1, a_2, \ldots, a_n)'$, $b = (b_1, b_2, \ldots, b_n)'$ and $[a, b] = \{x \in \mathbb{R}^n_+ \mid a \le x \le b\}$. For any $\boldsymbol{u} \in \mathcal{C}(\mathbb{R}_+)^n$ and $\underline{Q}, \overline{Q} \in \mathbb{R}^{n\times n}_+$ with $\underline{Q} \le \overline{Q}$ and $\langle \underline{Q}, \overline{Q}\rangle \ne \emptyset$, we define their product by

$$\langle \underline{Q}, \overline{Q}\rangle \boldsymbol{u} = \{Qu \mid Q \in \langle \underline{Q}, \overline{Q}\rangle, u \in \boldsymbol{u}\}. \tag{2.6}$$

Lemma 2.3 (Lemma 1.4 in [10]).

$$\langle \underline{Q}, \overline{Q}\rangle \boldsymbol{u} \in \mathcal{C}(\mathbb{R}_+)^n \quad \textit{for all } \boldsymbol{u} \in \mathcal{C}(\mathbb{R}_+)^n.$$

The following arithmetical notation is used in the sequel. Let $\widetilde{Q} = [\widetilde{q}_1, \widetilde{q}_2, \ldots, \widetilde{q}_n]'$ be a fuzzy set on $\mathcal{P}(S/S)$ with $\widetilde{q}_i \in \mathcal{F}([0,1])^n$ $(1 \le i \le n)$. Then, for $\widetilde{\boldsymbol{u}} = (\widetilde{u}_1, \widetilde{u}_2, \ldots, \widetilde{u}_n)' \in \mathcal{F}_c(\mathbb{R}_+)^n$, $\widetilde{Q}\widetilde{\boldsymbol{u}} \in \mathcal{F}(\mathbb{R}_+^n)$ is defined as follows:

$$(2.7) \qquad (\widetilde{Q}\widetilde{\boldsymbol{u}})(x) = \max_{\substack{x=Qu \\ Q\in\mathcal{P}(S/S), u\in\mathbb{R}_+^n}} \{\widetilde{Q}(Q) \wedge \widetilde{\boldsymbol{u}}(u)\}, \quad \text{for } x \in \mathbb{R}_+^n, \quad \text{where}$$

$$(2.8) \qquad \widetilde{\boldsymbol{u}}(u) = \min_{1\le i\le n} \{\widetilde{u}_i(u_i)\} \quad \text{with} \quad u = (u_1, u_2, \ldots, u_n) \in \mathbb{R}_+^n.$$

Lemma 2.4. *For any $\widetilde{\boldsymbol{u}} = (\widetilde{u}_1, \widetilde{u}_2, \ldots, \widetilde{u}_n)' \in \mathcal{F}_c(\mathbb{R}_+)^n$, we have:*

(i) $(\widetilde{Q}\widetilde{\boldsymbol{u}})_\alpha = \widetilde{Q}_\alpha \widetilde{\boldsymbol{u}}_\alpha$ *for* $\alpha \in [0,1]$;

(ii) $\widetilde{Q}\widetilde{\boldsymbol{u}} \in \mathcal{F}_c(\mathbb{R}_+)^n$.

Proof. By (2.7) we get $(\widetilde{Q}\widetilde{\boldsymbol{u}})_\alpha = \{Qu \mid q \in \widetilde{Q}_\alpha, u \in \widetilde{\boldsymbol{u}}_\alpha\}$. From (2.8) it holds $\widetilde{\boldsymbol{u}}_\alpha \in \mathcal{C}(\mathbb{R}_+)^n$, so that (i) follows by the definition (2.6). Also, (ii) follows obviously from Lemma 2.2 and 2.3. □

The addition and the scalar multiplication on $\mathcal{F}_c(\mathbb{R}_+)$ are defined as follows: For $\widetilde{a}, \widetilde{b} \in \mathcal{F}_c(\mathbb{R}_+)$ and $\lambda \in \mathbb{R}_+$, define

$$(\widetilde{a} + \widetilde{b})(x) := \sup_{\substack{x_1, x_2 \in \mathbb{R}_+ \\ x_1 + x_2 = x}} \{\widetilde{a}(x_1) \wedge \widetilde{b}(x_2)\},$$

$$\lambda\widetilde{a}(x) := \begin{cases} \widetilde{a}(x/\lambda) & \text{if } \lambda > 0 \\ I_{\{0\}}(x) & \text{if } \lambda = 0 \end{cases} \quad (x \in \mathbb{R}_+),$$

where I_A is the indicator of a set A. It is easily shown that, for $\alpha \in [0,1]$,

$$(\widetilde{a} + \widetilde{b})_\alpha = \widetilde{a}_\alpha + \widetilde{b}_\alpha \quad \text{and} \quad (\lambda\widetilde{a})_\alpha = \lambda\widetilde{a}_\alpha,$$

where the operation on sets is defined ordinary as $A + B := \{x + y \mid x \in A, y \in B\}$ and $\lambda A = \{\lambda x \mid x \in A\}$ for $A, B \subset \mathbb{R}_+$. The above operations are extended to those on $\mathcal{F}_c(\mathbb{R}_+)^n$ as follows: For $\widetilde{\boldsymbol{u}} = (\widetilde{u}_1, \widetilde{u}_2, \ldots, \widetilde{u}_n)'$, $\widetilde{\boldsymbol{v}} = (\widetilde{v}_1, \widetilde{v}_2, \ldots, \widetilde{v}_n)' \in \mathcal{F}_c(\mathbb{R}_+)^n$,

$$\widetilde{\boldsymbol{u}} + \widetilde{\boldsymbol{v}} = (\widetilde{u}_1 + \widetilde{v}_1, \widetilde{u}_2 + \widetilde{v}_2, \ldots, \widetilde{u}_n + \widetilde{v}_n)' \quad \text{and} \quad \lambda\widetilde{\boldsymbol{u}} = (\lambda\widetilde{u}_1, \lambda\widetilde{u}_2, \ldots, \lambda\widetilde{u}_n)'.$$

For $a = (a_1, a_2, \ldots, a_n)' \in \mathbb{R}_+^n$, $I_{\{a\}} = (I_{\{a_1\}}, I_{\{a_2\}}, \ldots, I_{\{a_n\}}) \in \mathcal{F}_c(\mathbb{R}_+)^n$ and $I_{\{a\}} + \widetilde{\boldsymbol{u}}$ is described simply by $a + \widetilde{\boldsymbol{u}}$. The Hausdorff metric on $\mathcal{C}(\mathbb{R}_+)$ is denoted by δ, i.e.,

$$\delta([a,b],[c,d]) := |a - c| \vee |b - d| \quad \text{for } [a,b],[c,d] \in \mathcal{C}(\mathbb{R}_+),$$

where $x \vee y = \max\{x, y\}$ for $x, y \in \mathbb{R}$. This metric can be extended to $\mathcal{F}_c(\mathbb{R}_+)^n$ by

$$\delta(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}}) = \max_{1\le i\le n} \sup_{\alpha\in[0,1]} \delta((\widetilde{u}_i)_\alpha, (\widetilde{v}_i)_\alpha)$$

for $\widetilde{\boldsymbol{u}} = (\widetilde{u}_1, \widetilde{u}_2, \ldots, \widetilde{u}_n)'$, $\widetilde{\boldsymbol{v}} = (\widetilde{v}_1, \widetilde{v}_2, \ldots, \widetilde{v}_n)' \in \mathcal{F}_c(\mathbb{R}_+)^n$. Then, it is known(c.f.[12]) that the metric space $(\mathcal{F}_c(\mathbb{R}_+)^n, \delta)$ is complete.

3. The fuzzy description of MDPs

In order to deal with the vague data and flexible requirements for nonhomogenuous MDPs we shall use a fuzzy set representation.

Let S and A be finite sets denoted by $S = \{1, 2, \ldots, n\}$ and $A = \{1, 2, \ldots, k\}$. Our sequential decision model consists of four objects:

$$(S, A, \{\widetilde{q}_{ij}(a) \in \mathcal{F}_c([0,1]),\ i, j \in S,\ a \in A\}, r),$$

where $r = r(i, a)$ is a function on $S \times A$ with $r \geq 0$. We interpret S as the set of states of some system and A as the set of actions available at each state. We denote by F the set of all functions from S to A. For any $f \in F$, we define the fuzzy set $\widetilde{Q}(f)$ on $\mathcal{P}(S/S)$ as follows:

$$\widetilde{Q}(f) := [\widetilde{q}_1(f), \widetilde{q}_2(f), \ldots, \widetilde{q}_n(f)]' \quad \text{where} \tag{3.1}$$

$$\widetilde{q}_i(f) := (\widetilde{q}_{i1}(f(i)), \widetilde{q}_{i2}(f(i)), \ldots, \widetilde{q}_{in}(f(i))) \quad (1 \leq i \leq n). \tag{3.2}$$

Note that the basic notations of (3.1) and (3.2) are defined in (2.1) and (2.2).

A policy π is a sequence $(f_1, f_2, \ldots)$ of functions with $f_t \in F$ $(t \geq 1)$. Let Π be the class of policies. We denote by f^∞ the policy $(h_1, h_2, \ldots)$ with $h_t = f$ for all $t \geq 1$ and some $f \in F$. Such a policy is called stationary and denoted simply by $f \in F$. The set of all stationary policies will be denoted by Π_F.

For any $f \in F$, let $r(f)$ be an n-dimensional column vector whose i-th element is $r(i, f(i))$. Applying Zadeh's extension principle(cf.[15]), the fuzzy expected total discounted reward up to time T from a policy π is a element of $\mathcal{F}(\mathbb{R}_+)^n$ and defined as follows:

$$\widetilde{\psi}_T(\pi) := (\widetilde{\psi}_T(1, \pi), \widetilde{\psi}_T(2, \pi), \ldots, \widetilde{\psi}_T(n, \pi))' \quad \text{and} \tag{3.3}$$

$$\widetilde{\psi}_T(i, \pi)(x) := \max\{\min_{1 \leq t \leq T} \widetilde{Q}(f_t)(Q_t)\} \quad \text{for all } x \in \mathbb{R}_+,\ 1 \leq i \leq n, \tag{3.4}$$

where the maximum is taken over

$$\begin{aligned}\{Q_1, Q_2, \ldots, Q_T \mid x = (r(f_1) + \beta Q_1 r(f_2) + \cdots + \beta^T Q_1 Q_2 \cdots Q_T r(f_{T+1}))_i,\\ Q_t \in \mathcal{P}(S/S)\ (1 \leq t \leq T)\}\end{aligned} \tag{3.5}$$

and β is a discounted factor with $0 < \beta < 1$.

Lemma 3.1 *For any policy $\pi \in \Pi$, we have:*
(i) $\widetilde{\psi}_T(\pi) \in \mathcal{F}_c(\mathbb{R}_+)^n$ *for all* $T \geq 1$;
(ii) $\{\widetilde{\psi}_T(\pi)\}$ *is a Caushy sequence.*

Proof. We show that, for example, (i) holds for $T = 2$. By (3.3) – (3.5),

$$\begin{aligned}&(\widetilde{\psi}_T(1, \pi)_\alpha, \widetilde{\psi}_T(2, \pi)_\alpha, \ldots, \widetilde{\psi}_T(n, \pi)_\alpha)' \\ &= \{r(f_1) + \beta Q_1 r(f_2) + \beta^2 Q_1 Q_2 r(f_3) \mid Q_i \in \widetilde{Q}(f_i)_\alpha,\ 1 \leq i \leq 2\} \\ &= r(f_1) + \beta \widetilde{Q}(f_1)_\alpha (r(f_2) + \beta \widetilde{Q}(f_2)_\alpha r(f_2)).\end{aligned}$$

Therefore, Lemma 2.2 and 2.3 it follows that

$$(\widetilde{\psi}_T(1,\pi)_\alpha, \widetilde{\psi}_T(2,\pi)_\alpha, \dots, \widetilde{\psi}_T(n,\pi)_\alpha)' \in \mathcal{C}(\mathbb{R}_+)^n,$$

which implies (i) for $T=2$. By the same method as the case of $T=2$, we can prove (i) for any T. Also, (ii) follows easily from the properties of the Hausdorff metric and the existence of the discount factor β $(0<\beta<1)$. □

By Lemma 3.1, we can define the infinite horizon fuzzy expected discounted reward(FEDR) from a policy π by

$$\widetilde{\psi}(\pi) := \lim_{T\to\infty} \widetilde{\psi}_T(\pi).$$

Here, we will give a partial order $\preccurlyeq$ on $\mathcal{C}(\mathbb{R}_+)$ by the definition: For $[a,b]$, $[c,d] \in \mathcal{C}(\mathbb{R}_+)$,

$$\begin{aligned} &[a,b] \preccurlyeq [c,d] \quad \text{if} \quad a \le c \text{ and } b \le d, \\ &[a,b] \prec [c,d] \quad \text{if} \quad [a,b] \preccurlyeq [c,d] \text{ and } [a,b] \ne [c,d]. \end{aligned}$$

This partial order $\preccurlyeq$ on $\mathcal{C}(\mathbb{R}_+)$ is extended to that of $\mathcal{F}_c(\mathbb{R}_+)$, called a fuzzy max order, as follows: For $\widetilde{u}$, $\widetilde{v} \in \mathcal{F}_c(\mathbb{R}_+)$,

$$\begin{aligned} &\widetilde{u} \preccurlyeq \widetilde{v} \quad \text{if} \quad \widetilde{u}_\alpha \preccurlyeq \widetilde{v}_\alpha \text{ for all } \alpha \in [0,1], \\ &\widetilde{u} \prec \widetilde{v} \quad \text{if} \quad \widetilde{u} \preccurlyeq \widetilde{v} \text{ and } \widetilde{u} \ne \widetilde{v}. \end{aligned}$$

Also, as a further extension, the partial order on $\mathcal{F}_c(\mathbb{R}_+)^n$ is given by the definition: For $\widetilde{\boldsymbol{u}} = (\widetilde{u}_1, \widetilde{u}_2, \dots, \widetilde{u}_n)'$, $\widetilde{\boldsymbol{v}} = (\widetilde{v}_1, \widetilde{v}_2, \dots, \widetilde{v}_n)' \in \mathcal{F}_c(\mathbb{R}_+)^n$,

$$\begin{aligned} &\widetilde{\boldsymbol{u}} \preccurlyeq \widetilde{\boldsymbol{v}} \quad \text{if} \quad \widetilde{u}_i \preccurlyeq \widetilde{v}_i \text{ for all } i = 1,2,\dots,n, \\ &\widetilde{\boldsymbol{u}} \prec \widetilde{\boldsymbol{v}} \quad \text{if} \quad \widetilde{\boldsymbol{u}} \preccurlyeq \widetilde{\boldsymbol{v}} \text{ and } \widetilde{\boldsymbol{u}} \ne \widetilde{\boldsymbol{v}}. \end{aligned}$$

Our problem is to maximize the $\widetilde{\psi}(\pi)$ over all $\pi \in \Pi$ with respect to the partial order $\preccurlyeq$.

The following lemma is used in the sequel whose proof is easily done.

Lemma 3.2 *Let a sequence* $\{\widetilde{\boldsymbol{u}}_n\} \subset \mathcal{F}_c(\mathbb{R}_+)^n$ *be such that* $\widetilde{\boldsymbol{u}}_1 \preccurlyeq \widetilde{\boldsymbol{u}}_2 \preccurlyeq \cdots$, *and* $\lim_{k\to\infty} \widetilde{\boldsymbol{u}}_k = \widetilde{\boldsymbol{u}}$ *for some* $\widetilde{\boldsymbol{u}} \in \mathcal{F}_c(\mathbb{R}_+)^n$. *Then, it holds that* $\widetilde{\boldsymbol{u}}_1 \preccurlyeq \widetilde{\boldsymbol{u}}$.

4. Stationary policies and operators

In this section, the infinite horizon FEDR from a stationary policy is given as a unique fixed point of a corresponding operator. Associated with each function $f \in F$ is a corresponding operator $U(f) : \mathcal{F}_c(\mathbb{R}_+)^n \to \mathcal{F}_c(\mathbb{R}_+)^n$ defined as follows: For $\widetilde{\boldsymbol{u}} \in \mathcal{F}_c(\mathbb{R}_+)^n$,

(4.1) $$U_f \widetilde{\boldsymbol{u}} = r(f) + \beta \widetilde{Q}(f) \widetilde{\boldsymbol{u}},$$

where the arithmetics in (4.1) are defined in (2.7). Note that from Lemma 1.4 U_f is well-defined.

For any policy $\pi = (f_1, f_2, \dots)$, let $\pi^{-l} = (f_{l+1}, f_{l+2}, \dots)$ for each $l \ge 1$. The sequence $\{\widetilde{\psi}_T(\pi)\}_{T=1}^\infty$ is recursively described.

Lemma 4.1 *For any policy* $\pi = (f_1, f_2, \dots)$, *we have*

(4.2) $$\widetilde{\psi}_T(\pi) = U_{f_1} U_{f_2} \cdots U_{f_l} \widetilde{\psi}_{T-l}(\pi^{-l}) \quad \text{for each} \quad l \ge 1.$$

Proof. From (3.3)–(3.5), we get, for each $\alpha \in [0,1]$,

$$\widetilde{\psi}_2(i,\pi)_\alpha = (r(f_1) + \beta\widetilde{Q}(f_1)r(f_2))_\alpha = r(f_1) + \beta\widetilde{Q}(f_1)_\alpha r(f_2), \text{ from Lemma 2.4 (i),}$$

Since $\widetilde{\psi}_1(\pi^{-1}) = r(f_2)$, (4.2) holds for $T = 2$ and $l = 1$. By induction on T and k, we can easily proved (4.2). □

Lemma 4.2. *Let $f \in F$. Then we have:*

(i) *U_f is a contraction with modulus β, i.e.,*
$\delta(U_f\widetilde{\boldsymbol{u}}, U_f\widetilde{\boldsymbol{v}}) \le \beta\delta(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}})$ for $\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}} \in \mathcal{F}_c(\mathbb{R}_+)^n$,

(ii) *U_f is monotone, i.e., $\widetilde{\boldsymbol{u}} \preccurlyeq \widetilde{\boldsymbol{v}}$ implies $U_f\widetilde{\boldsymbol{u}} \preccurlyeq U_f\widetilde{\boldsymbol{v}}$.*

Proof. For any $\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}} \in \mathcal{F}_c(\mathbb{R}_+)^n$, from the property of the Hausdorff metric, it holds $\delta(U_f\widetilde{\boldsymbol{u}}, U_f\widetilde{\boldsymbol{v}}) \le \beta\delta(\widetilde{Q}(f)\widetilde{\boldsymbol{u}}, \widetilde{Q}(f)\widetilde{\boldsymbol{v}})$. Using Lemma 2.4 (i), we get

$$\delta((\widetilde{Q}(f)\widetilde{\boldsymbol{u}})_\alpha, (\widetilde{Q}(f)\widetilde{\boldsymbol{v}})_\alpha) = \delta(\widetilde{Q}(f)_\alpha\widetilde{\boldsymbol{u}}_\alpha, \widetilde{Q}(f)_\alpha\widetilde{\boldsymbol{v}}_\alpha) \le \delta(\widetilde{\boldsymbol{u}}_\alpha, \widetilde{\boldsymbol{v}}_\alpha) \ (0 \le \alpha \le 1).$$

So, we have $\delta(U_f\widetilde{\boldsymbol{u}}, U_f\widetilde{\boldsymbol{v}}) \le \beta\delta(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}})$, which implies (i). Also, (ii) follows obviously. □

By Lemma 3.1, $\widetilde{\psi}_T(f) = U_f\widetilde{\psi}_{T-1}(f)$ for all $T \ge 2$. As $T \to \infty$ in the above, $\widetilde{\psi}(f)$ is a fixed point of U_f. Thus, the following characterization of $\widetilde{\psi}(f)$, formulated as a theorem, are immediate.

Theorem 4.1. *For any $f \in F$, $\widetilde{\psi}(f)$ is a unique solution of the following fuzzy equation:*

$$\widetilde{\boldsymbol{u}} = U_f\widetilde{\boldsymbol{u}}, \quad \widetilde{\boldsymbol{u}} \in \mathcal{F}_c(\mathbb{R}_+)^n. \tag{4.3}$$

Applying Lemma 2.4 (i), (4.3) can be rewritten by the following α-cut interval equation:

$$\widetilde{\boldsymbol{u}}_\alpha = r(f) + \beta\widetilde{Q}(f)_\alpha\widetilde{\boldsymbol{u}}_\alpha, \ 0 \le \alpha \le 1, \tag{4.4}$$

where $\widetilde{\boldsymbol{u}}_\alpha = ((\widetilde{u}_1)_\alpha, (\widetilde{u}_2)_\alpha, \ldots, (\widetilde{u}_n)_\alpha)' \in \mathcal{C}(\mathbb{R}_+)^n$ and $\widetilde{Q}(f)_\alpha = \langle \underline{Q}_\alpha, \overline{Q}_\alpha \rangle$ with $\underline{Q}_\alpha \le \overline{Q}_\alpha$.

By a contraction of U_f, the following holds.

Corollary 4.1. *For any $f \in F$ and $\widetilde{\boldsymbol{u}} \in \mathcal{F}_c(\mathbb{R}_+)^n$, $\widetilde{\psi}(f) = \lim_{l\to\infty} U_f^l\widetilde{\boldsymbol{u}}$.*

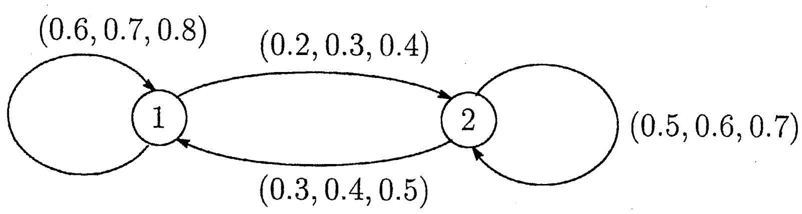
As a simple example, we consider a fuzzy treatment for a machine maintenance problem dealt with in ([13], p.1, p.17–18).

An example (a machine maintenance problem). A machine can be operated synchronously, say, once an hour. At each period there are two states; one is operating(state 1), and the other is in failure(state 2). If the machine fails, it can be restored to perfect functioning by repair. At each period, if the machine is running, we earn the return of $ 3.00 per period; the fuzzy set of probability of being in state 1 at the next step is (0.6, 0.7, 0.8) and that of the probability of moving to state 2 is (0.2, 0.3, 0.4), where for any $0 \le a < b < c \le 1$, the fuzzy number (a,b,c) on $[0,1]$ is defined by

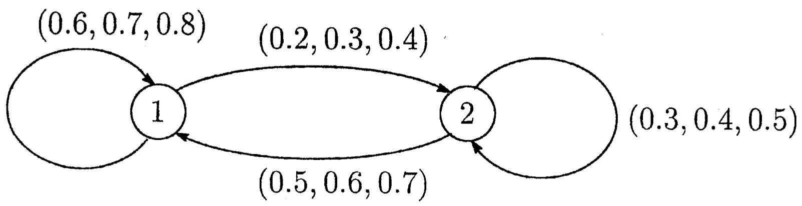
$$(a,b,c)(x) = \begin{cases} (x-a)/(b-a) \vee 0 & \text{if } 0 \le x \le b, \\ (x-c)/(b-c) \vee 0 & \text{if } b \le x \le 1. \end{cases}$$

If the machine is in failure, we have two actions to repair the failed machine; one is a usual repair, denoted by 1, that yields the cost of \$ 1.00(that is, a return of $-\$1.00$) with the fuzzy set (0.3, 0.4, 0.5) of the probability moving in state 1 and the fuzzy set (0.5, 0.6, 0.7) of the probability being in state 2; another is a rapid repair, denoted by 2, that requires the cost of \$2.00(that is, a return of $-\$2.00$) with the fuzzy set (0.5, 0.6, 0.7) of the probability moving in state 1 and the fuzzy set (0.3, 0.4, 0.5) of the probability being in state 2.

For the model considered, $S = \{1, 2\}$ and there exists two stationary policies, $F = \{f_1, f_2\}$ with $f_1(2) = 1$ and $f_2(2) = 2$, where f_1 denotes a policy of the usual repair and f_2 a policy of the rapid repair. The state transition diagrams of two policies are shown in Figure 1.



(a) Usual repair f_1



(b) Rapid repair f_2

Figure.1 Transition diagrams.

We easily observe that

$$r(f_1) = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \quad \text{and} \quad \tilde{Q}(f_1) = \begin{pmatrix} (0.6, 0.7, 0.8) & (0.2, 0.3, 0.4) \\ (0.3, 0.4, 0.5) & (0.5, 0.6, 0.7) \end{pmatrix},$$

Now, applying Theorem 4.1, we can obtain the infinite horizon FEDR as a unique solution of (4.4). We observe that

$$\tilde{Q}(f_1)_\alpha = \left\langle \begin{pmatrix} 0.6 + 0.1\alpha & 0.2 + 0.1\alpha \\ 0.3 + 0.1\alpha & 0.5 + 0.1\alpha \end{pmatrix}, \begin{pmatrix} 0.8 - 0.1\alpha & 0.4 - 0.1\alpha \\ 0.5 - 0.1\alpha & 0.7 - 0.1\alpha \end{pmatrix} \right\rangle,$$

which is a convex hull of

$$\left\{ \begin{matrix} \begin{pmatrix} 0.6 + 0.1\alpha & 0.4 - 0.1\alpha \\ 0.3 + 0.1\alpha & 0.7 - 0.1\alpha \end{pmatrix}, \begin{pmatrix} 0.8 - 0.1\alpha & 0.2 + 0.1\alpha \\ 0.5 - 0.1\alpha & 0.5 + 0.1\alpha \end{pmatrix} \\ \begin{pmatrix} 0.6 + 0.1\alpha & 0.4 - 0.1\alpha \\ 0.5 - 0.1\alpha & 0.5 + 0.1\alpha \end{pmatrix}, \begin{pmatrix} 0.8 - 0.1\alpha & 0.2 + 0.1\alpha \\ 0.3 + 0.1\alpha & 0.7 - 0.1\alpha \end{pmatrix} \end{matrix} \right\}.$$

So, putting $\widetilde{\psi}(f_1)_\alpha = ([x_1^\alpha, y_1^\alpha]', [x_2^\alpha, y_2^\alpha]')$, the α-cut interval equations (4.4) with $\beta = 0.9$ become:

$$x_1^\alpha = 3 + 0.9\{(0.6x_1^\alpha + 0.4x_2^\alpha + 0.1\alpha(x_1^\alpha - x_2^\alpha)) \wedge (0.8x_1^\alpha + 0.2x_2^\alpha + 0.1\alpha(-x_1^\alpha + x_2^\alpha))\},$$

$$y_1^\alpha = 3 + 0.9\{(0.6y_1^\alpha + 0.4y_2^\alpha + 0.1\alpha(y_1^\alpha - y_2^\alpha)) \vee (0.8y_1^\alpha + 0.2y_2^\alpha + 0.1\alpha(-y_1^\alpha + y_2^\alpha))\},$$

$$x_2^\alpha = -1 + 0.9\{(0.5x_1^\alpha + 0.5x_2^\alpha + 0.1\alpha(x_2^\alpha - x_1^\alpha)) \wedge (0.3x_1^\alpha + 0.7x_2^\alpha + 0.1\alpha(x_1^\alpha - x_2^\alpha))\},$$

$$y_2^\alpha = -1 + 0.9\{(0.5y_1^\alpha + 0.5y_2^\alpha + 0.1\alpha(y_2^\alpha - y_1^\alpha)) \vee (0.3y_1^\alpha + 0.7y_2^\alpha + 0.1\alpha(-y_1^\alpha + y_2^\alpha))\}.$$

After a simple calculation, we find

$$\widetilde{\psi}(f_1)_\alpha = \Big([\frac{750 + 360\alpha}{73}, \frac{1470 - 360\alpha}{73}], [\frac{1350 + 360\alpha}{73}, \frac{1070 - 360\alpha}{73}]\Big)',$$

which leads to

$$\widetilde{\psi}(f_1) = \Big((\frac{750}{73}, \frac{1110}{73}, \frac{1470}{73}), (\frac{350}{73}, \frac{710}{73}, \frac{1070}{73})\Big)'.$$

5. Pareto optimality

Here, we confine our attention to the class of stationary policies, which simplifies our discussion in the sequel. A policy $f^* \in \Pi_F$ is called Pareto optimal if there does not exist $f \in \Pi_F$ such that $\widetilde{\psi}(f^*) \prec \widetilde{\psi}(f)$. In this section, we derive the optimal equation, by which Pareto optimal policies are characterized.

The following important result is crucial to the development in the characterization of Pareto optimality.

Lemma 5.1. *For any $f, g \in F$, suppose that*

$$\widetilde{\psi}(f) \ \begin{Bmatrix}\preccurlyeq \\ \prec\end{Bmatrix} \ U_g\widetilde{\psi}(f). \tag{5.1}$$

Then, it holds that

$$\widetilde{\psi}(f) \ \begin{Bmatrix}\preccurlyeq \\ \prec\end{Bmatrix} \ \widetilde{\psi}(g). \tag{5.2}$$

Proof. Suppose that $\widetilde{\psi}(f) \ \begin{Bmatrix}\preccurlyeq \\ \prec\end{Bmatrix} \ U_g\widetilde{\psi}(f)$. Then, we have from Lemma 4.2 (ii) that

$$\widetilde{\psi}(f) \ \begin{Bmatrix}\preccurlyeq \\ \prec\end{Bmatrix} \ U_g\widetilde{\psi}(f) \ \preccurlyeq \ U_g^l\widetilde{\psi}(f) \ \ (l \geq 2),$$

So, taking the limit in the above as $l \to \infty$, (5.2) follows from Lemma 3.2. $\square$

Let $\boldsymbol{D}$ be an arbitrary subset of $\mathcal{F}_c(\mathbb{R}_+)^n$. A point $\widetilde{\boldsymbol{u}} \in \boldsymbol{D}$ is called an efficient element of $\boldsymbol{D}$ with respect to $\preccurlyeq$ on $\mathcal{F}_c(\mathbb{R}_+)^n$ if and only if it holds that there does not exist $\widetilde{\boldsymbol{v}} \in \boldsymbol{D}$ such that $\widetilde{\boldsymbol{u}} \prec \widetilde{\boldsymbol{v}}$. We denote by $\mathrm{eff}(\boldsymbol{D})$ the set of all elements of $\boldsymbol{D}$ efficient with respect to $\preccurlyeq$ on $\mathcal{F}_c(\mathbb{R}_+)^n$. For any $\widetilde{\boldsymbol{u}} \in \mathcal{F}_c(\mathbb{R}_+)^n$, let $\boldsymbol{U}(\widetilde{\boldsymbol{u}}) := \mathrm{eff}(\{U_f\widetilde{\boldsymbol{u}} \mid f \in F\})$. Note that $\boldsymbol{U}(\widetilde{\boldsymbol{u}}) \subset \mathcal{F}_c(\mathbb{R}_+)^n$.

Here, we consider the following fuzzy equation including efficient set-functions $\boldsymbol{U}(\cdot)$ on $\mathcal{F}_c(\mathbb{R}_+)^n$:

$$\widetilde{\boldsymbol{u}} \in \boldsymbol{U}(\widetilde{\boldsymbol{u}}), \quad \widetilde{\boldsymbol{u}} \in \mathcal{F}_c(\mathbb{R}_+)^n. \tag{5.3}$$

The equation (5.3) is called an optimality equation, by which Pareto optimal policies are characterized. A solution of (5.3), $\widetilde{\boldsymbol{u}}$, is called maximal if there does not exist any solution $\widetilde{\boldsymbol{u}}'$ of (5.3) such that $\widetilde{\boldsymbol{u}} \prec \widetilde{\boldsymbol{u}}'$. Pareto optimal policies are characterized by maximal solutions of the optimality equation (5.3).

Theorem 5.1. *A policy f is Pareto optimal if and only if a fixed point of the corresponding U_f, $\widetilde{\psi}(f)$, is a maximal solution to the optimal equation (5.3).*

Proof. The proof of "only if "part is easily obtained from Lemma 5.1. In order to prove "if "part, suppose that $\widetilde{\psi}(f)$ is a maximal solution of (5.3) but f is not Pareto optimal. Then, there exists $f^{(1)} \in F$ such that $\widetilde{\psi}(f) \prec \widetilde{\psi}(f^{(1)})$.

Now, suppose that $\widetilde{\psi}(f^{(1)}) \notin \text{eff}(\widetilde{\psi}(f^{(1)}))$. This assumption assures that there exists $f^{(2)} \in F$ satisfying $\widetilde{\psi}(f^{(1)}) \prec U_{f^{(2)}}\widetilde{\psi}(f^{(1)})$, which implies from (5.1) that $\widetilde{\psi}(f^{(1)}) \prec \widetilde{\psi}(f^{(2)})$. By repeating this method successively, we come to the conclusion that there exists $f^{(l)} \in F$ such that $\widetilde{\psi}(f) \prec \widetilde{\psi}(f^{(l)})$ and $\widetilde{\psi}(f^{(l)})$ satisfies (5.3), which contradicts that $\widetilde{\psi}(f)$ is maximal, as required. □

Remark. For vector-valued discounted MDPs, Furukawa[2] and White[17] had derived the optimality equation including efficient set-function on $\mathbb{R}^n$, by that Pareto optimal policies are characterized. The form of the optimal equation (5.3) is corresponding to a fuzzy version of MDPs.

For the machine maintenance problem given in Section 4, we find that

$$U_{f_2}\widetilde{\psi}(f_1) = \Big(\big(\frac{750}{73}, \frac{1110}{73}, \frac{1470}{73}\big), \big(\frac{349}{73}, \frac{709}{73}, \frac{1069}{73}\big) \Big)',$$

Recall that

$$U_{f_1}\widetilde{\psi}(f_1) = \widetilde{\psi}(f_1) = \Big(\big(\frac{750}{73}, \frac{1110}{73}, \frac{1470}{73}\big), \big(\frac{350}{73}, \frac{710}{73}, \frac{1070}{73}\big) \Big)',$$

which satisfies $U_{f_2}\widetilde{\psi}(f_1) \prec \widetilde{\psi}(f_1)$. Thus, $\widetilde{\psi}(f_1) \in \text{eff}(\{U_f\widetilde{\psi}(f_1) \mid f \in F)$, so that from Theorem 5.1 f_1 is Pareto optimal. In fact, we can find, by solving (4.4) for f_2, that

$$\widetilde{\psi}(f_2) = \Big(\big(\frac{930}{91}, \frac{1380}{91}, \frac{1830}{91}\big), \big(\frac{430}{91}, \frac{880}{91}, \frac{1330}{91}\big) \Big)', \text{ and } \widetilde{\psi}(f_2) \prec \widetilde{\psi}(f_1).$$

References

[1] Blackwell, D., Discrete dynamic programming, *Ann. Math. Stat.* **33** (1962), 719-726.

[2] Furukawa, N., Characterization of optimal policies in vector-valued Markovian decision processes, *Math. Oper. Res.* **5** (1980), 271-279.

[3] Hartfiel, D. J. and Seneta, E., On the theory of Markov Set-chains, *Adv. Appl. Prob.* **26** (1994), 947-964.

[4] Hartfiel, D. J., *Markov Set-chains*, (1998), Springer-Verlag, Berlin.

[5] Hinderer, K., *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter,* (1970), Springer-Verlag, New York.

[6] Hosaka, M. and Kurano, M., Non-discounted Optimal policy in controlled Markov Set-chains, *To appear in Journal of Opern. Res. of Japan*, **42** (1999).

[7] Howard, R., *Dynamic Programming and Markov processes*, (1960), MIT Press, Cambridge MA.

[8] Kruse, R., Buck-emden, R. and Cordes, R., Processor power considerations - an application to fuzzy Markov chains, *Fuzzy Sets and Systems*, **21** (1987), 289-299.

[9] Kurano, M., Yasuda, M., Nakagami, J. and Yoshida, Y., Markov-type fuzzy decision processes with a discounted reward on a closed interval, *European J. O. R.*, **92** (1996), 649-662.

[10] Kurano, M., Song, J., Hosaka, M. and Huang, Y., Controlled Markov Set-Chains with Discounting, *J. Appl. Prob.*, **35** (1998), 293-302.

[11] Kurano, M., Yasuda, M. and Nakagami, J., Interval methods for uncertain Markov decision processes, *submitted to the International Workshop on Markov Processes and Controlled Markov Chains*, Changsha, Husan, China on August 22-28, 1999.

[12] Kuratowski, K, *Topology*, (1966), Academic Press, New York.

[13] Mine, H. and Osaki, S., *Markov Decision Processes*, (1970), Elsevier, Amsterdam.

[14] Nenmaier, A., New techniques for the analyses of linear interval equations, *Linear Algebra Appli.*, **58** (1984), 273-325.

[15] Novák, V., *Fuzzy sets and their applications*, (1989), Adam Hilger, Bristole-Boston.

[16] Puterman, M. L., *Markov decision processes: Discrete Stochastic Dynamic Programming*, (1994), John Wiley & Sons, INC.

[17] White, D. J., Multi-objective infinite-horizon discounted Markov Decision Processes, *J. Math. Anal. Appl.*, **89** (1982), 639-647.

[18] Yoshida, Y., A time-average fuzzy reward criterion in fuzzy decision processes, *Information Sciences*, **110** (1998), 103-112.

[19] Zadeh, L. A., Fuzzy sets, *Inform. and Control*, **8** (1965), 338-353.